

METHOD AND SYSTEM FOR DETECTING DEVIATIONS IN DATA TABLES

ABSTRACT OF THE DISCLOSURE

5 A method and system for automatically detecting deviations in a data table comprising a multitude
of records and a multitude of columns. A column of the data table is selected as a classification
column and a classification tree is calculated with respect to the classification column. Each edge
of the classification tree is associated with a predicate. The leaf nodes of the classification tree are
associated with a leaf record set comprising the subset of records for which the class predicate
10 comprising all predicates along a path from a root node of the classification tree to the leaf nodes
evaluates to TRUE. Leaf nodes are associated with a leaf label representing an expected value in
the classification column for the corresponding leaf record sets. From the leaf record sets all
records deviating with respect to the corresponding classification column from the leaf label are
determined as deviation sets.